Fouille de données et Apprentissage statistique

Frédéric Precioso

PRÉSENTATION ESP
LABORATOIRE I3S – **EQUIPE KEIA**

F. PRECIOSO

31/05/2013

Fouille de données et Apprentissage statistique

Laboratoire I3S - Equipe KEIA

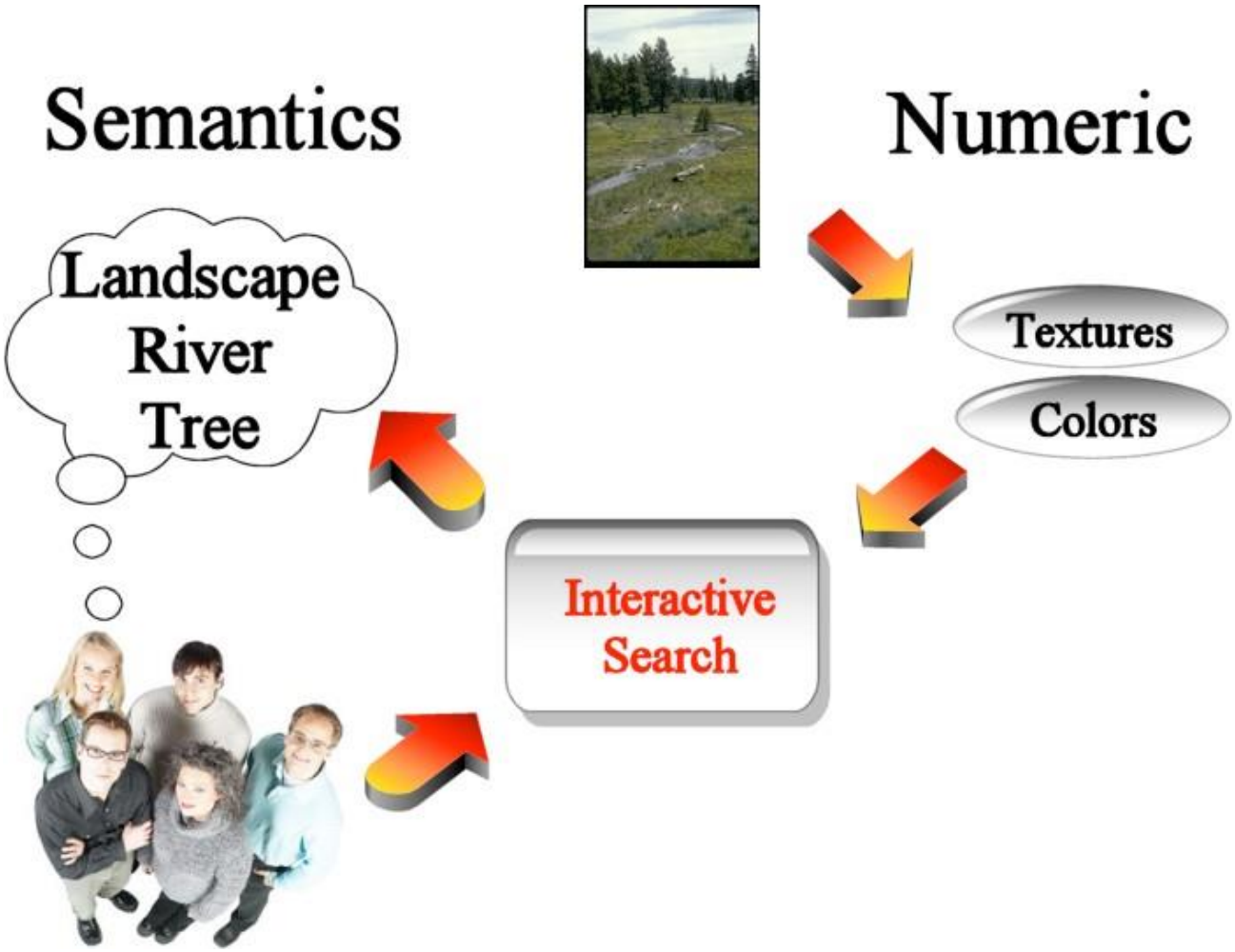
- Enseignants-chercheurs
  - Nicolas Pasquier
  - Célia da Costa Pereira
  - Christel Dartigues-Pallez
  - Denis Pallez
  - Frédéric Precioso
- Doctorants
  - ***Kartick Chandra MONDAL***, EMMA (Erasmus Mundus Mobility for Asia), Data mining algorithms for bioinformatics, ***2009-2013***
  - ***Somsack INTHASONE***, EMMA (Erasmus Mundus Mobility for Asia), Knowledge extraction techniques for biodiversity studies, ***2012-2015***
  - ***Romaric Pighetti***, MESR, Recherche d'information par le contenu en combinant algorithmes évolutionnaires et SVM, ***2012-2015***

Activités

- Problèmes
  - Clustering : regroupement de données “similaires”
  - Classification : construction de modèles explicites ou implicites de classes
  - Associations : relations de co-occurrence
  - Décision pour des données dynamiques et Planning
  - Apprentissage Interactif
- Méthodes
  - Algorithmes Evolutionaires
  - Machines à Vecteurs de Support
  - Systèmes Multi-Agents
  - Boosting
  - Réseaux de Neurones artificiels
  - Arbres de décision et Forêts aléatoires
  - Treillis de Galois et closure
  - Inférence Bayésienne Naïve
  - …

Information retrieval from The Big Data

CBIR systems: bridging the "*semantic gap*"

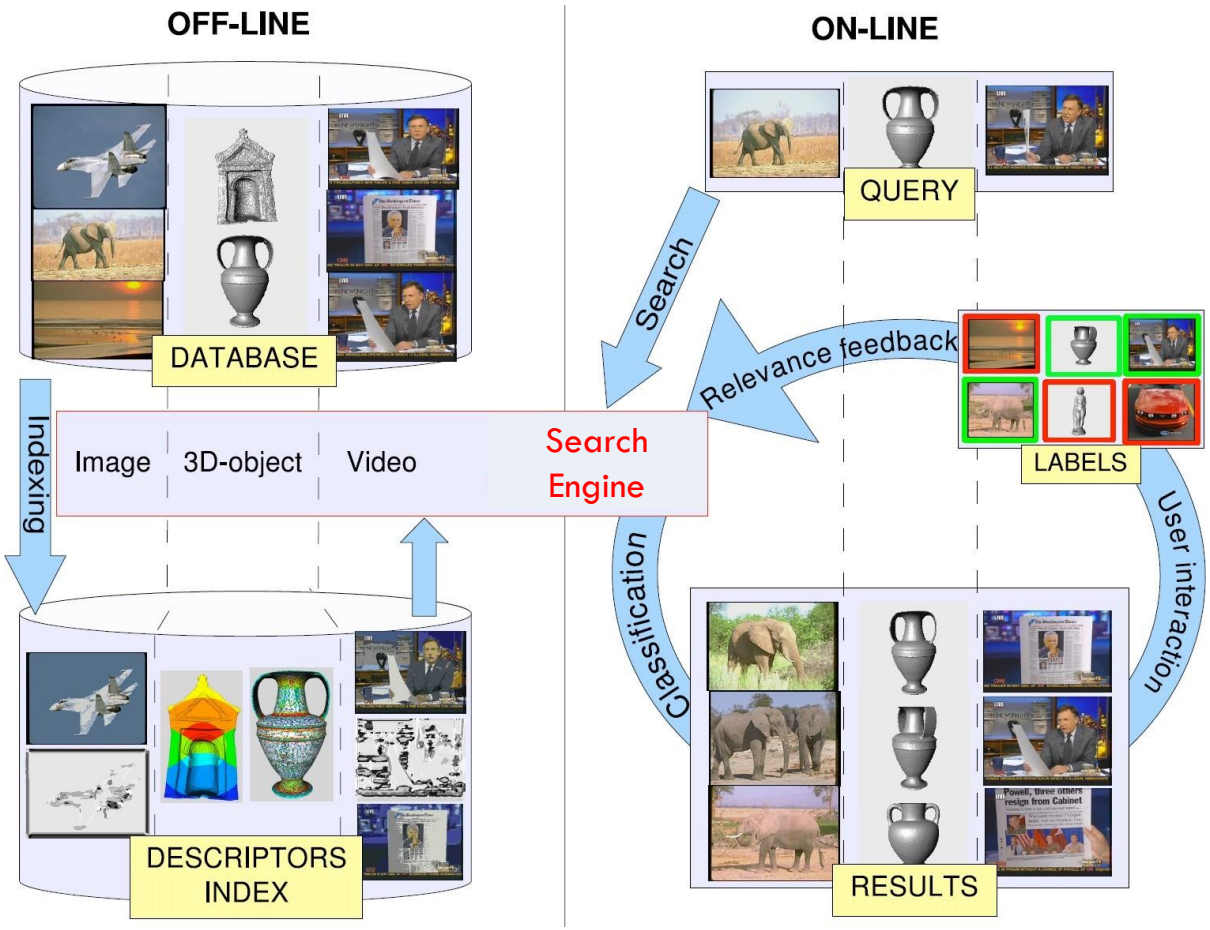


Information mining in The Big Data

- Increasing multimedia information stream…
- Mega-databases:
  - **YouTube** = May 2010, more videos uploaded in 60 days than filmed by the 3 biggest US majors in 60 years [1]; 20 h of video uploaded per minute in May 2009, 24 h per minute in March 2010, to reach ***1 h per minute*** few months ago [1, 4]
  - **Flickr** = 3000 images uploaded every minute to reach 5 billions in September 2010 [1]
- The Digital Universe was about 281 ExaBytes (281 Billions of GigaBytes) in 2007; it increased 10 times since then [2]
- Images and videos, captured by several billions of devices (cameras, webcams, phones) represent the biggest part of this incredibly big amount of data [3]

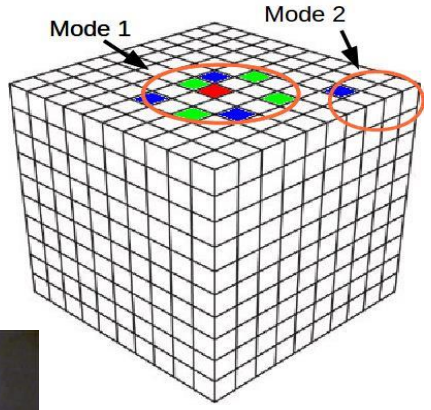
- *[1] website-monitoring.com*
- *[2] "digital universe.", http://eon.businesswire.com/releases/information/digital/prweb509640.htm*
- *[3] "major source." , http ://www.emc.com/collateral/analyst-reports/diverse-exploding-digitaluniverse.pdf*
- *[4] http://www.onehourpersecond.com/*

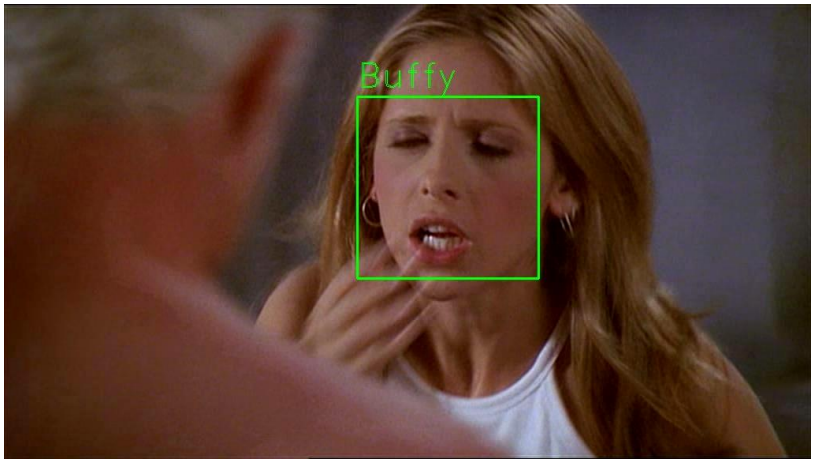
CBIR systems



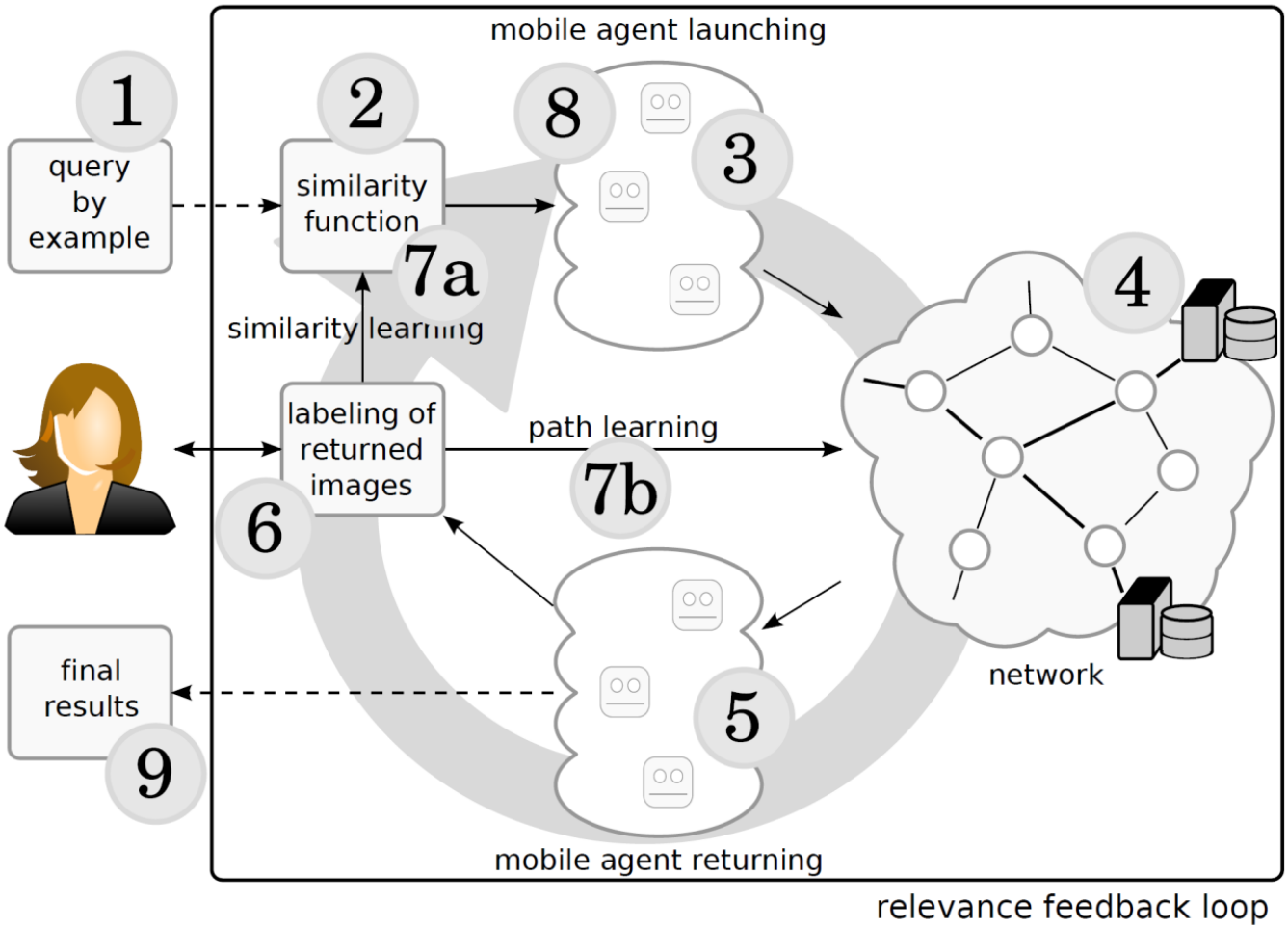
Information mining in The Big Data

- Text Mining in short messages (tweets, SMS,…)
- Human action recognition
- Large-scale image and video retrieval





Information mining in The Big Data



Credits Dr. D. Picard

e-Health

Health-Autonomy

VividIA

Virtu4D

Personne en mouvement

Localisation 2D dans l'image

Représentation dans l'environnement virtuel

Estimation de pose 3D

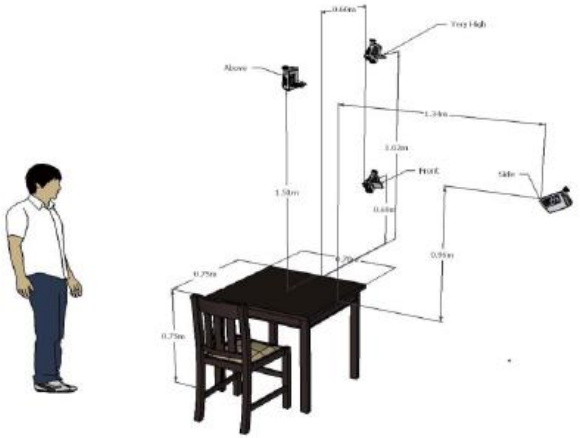


Fig 3: Setup - Exact placement of cameras for recording

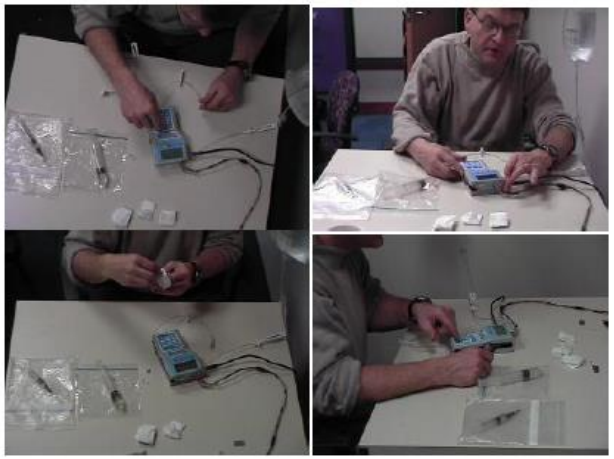
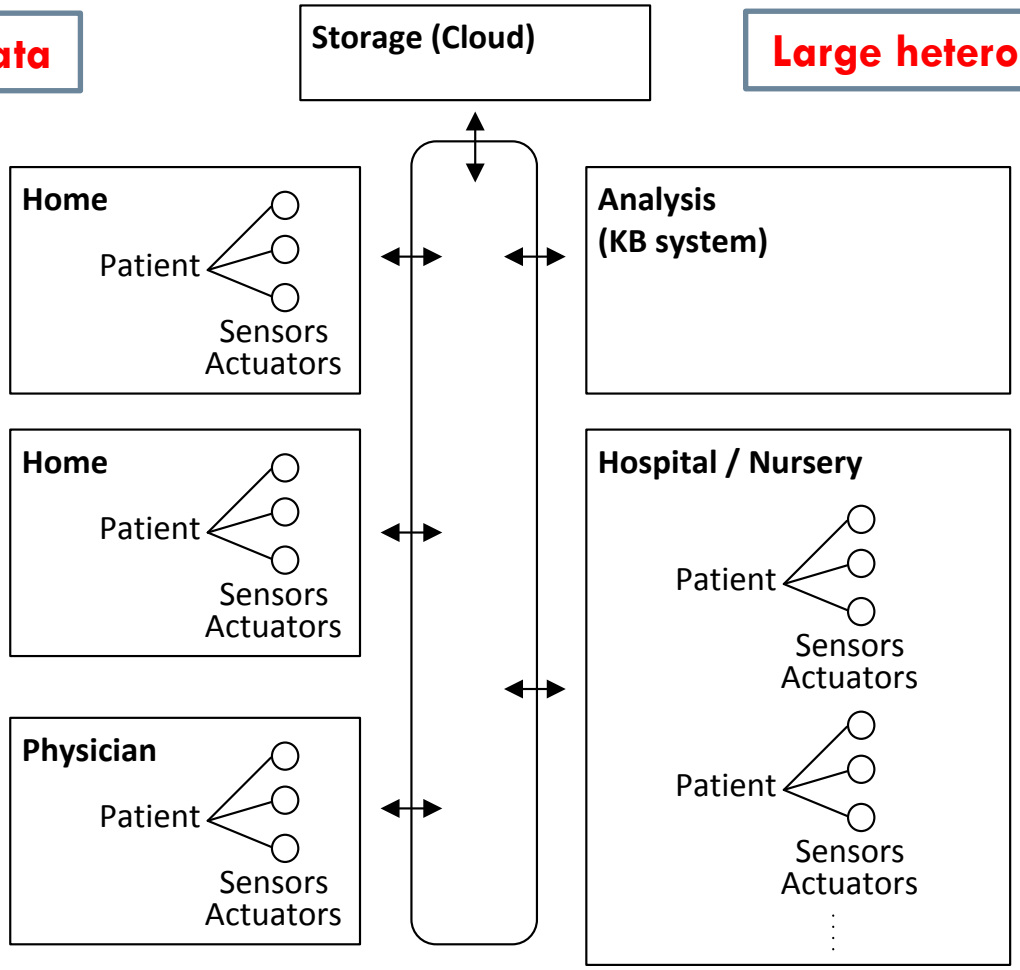


Fig 4: Example of videos from the different cameras. From left to right and top to bottom, images are from the "above" camera, the "front" camera, the "very high" camera and the "side" camera

Health 2.0

Patient/context data

Large heterogeneous sensor data



Ontologies describing application-specific knowledge

Laboratoire I3S - Equipe KEIA

- Nicolas Pasquier
- Célia da Costa Pereira
- Christel Dartigues-Pallez
- Denis Pallez
- *Frédéric Precioso*

Merci !

Avez-vous des questions ?